



# Repetitive Architecture of the *Haemophilus influenzae* Hia Trimeric Autotransporter

Guoyu Meng<sup>1</sup>, Joseph W. St. Geme III<sup>2</sup> and Gabriel Waksman<sup>1\*</sup>

<sup>1</sup>*Institute of Structural and Molecular Biology at UCL/ Birkbeck, Malet Street, London WE1C 7HX, UK*

<sup>2</sup>*Departments of Pediatrics and Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC 27710, USA*

Received 19 August 2008;  
received in revised form  
30 September 2008;  
accepted 30 September 2008

Available online  
11 October 2008

Edited by J. Bowie

The Hia autotransporter of *Haemophilus influenzae* belongs to the trimeric autotransporter subfamily and mediates bacterial adherence to the respiratory epithelium. In this report, we show that the structure of Hia is characterized by a modular architecture containing repeats of structurally distinct domains. Comparison of the structures of HiaBD1 and HiaBD2 adhesive repeats and a nonadhesive repeat (a novel fold) shed light on the structural determinants of Hia adhesive function. Examination of the structure of an extended version of the Hia translocator domain revealed the structural transition between the C-terminal translocator domain and the N-terminal passenger domain, highlighting a highly intertwined domain that is ubiquitous among trimeric autotransporters. Overall, this study provides important insights into the mechanism of Hia adhesive activity and the overall structure of trimeric autotransporters.

© 2008 Elsevier Ltd. Open access under [CC BY license](#).

**Keywords:** Hia; autotransporter; adhesin; structural biology; *Haemophilus influenzae*

## Introduction

*Haemophilus influenzae* is a Gram-negative bacterium that causes disease exclusively in humans. Nonencapsulated (nontypeable) strains are responsible for most *H. influenzae* localized respiratory tract diseases such as otitis media, sinusitis, and pneumonia, and encapsulated strains account for most *H. influenzae* invasive diseases, including sepsis and meningitis.<sup>1</sup>

The pathogenesis of *H. influenzae* disease begins with colonization of the upper respiratory tract,<sup>2,3</sup> followed by either contiguous spread within the respiratory tract or invasion of the bloodstream. The process of colonization requires bacterial adherence to the respiratory epithelium and is facilitated by both pilus and nonpilus adhesive factors.<sup>4</sup> In a subset of nontypeable strains, the major nonpilus adhesin is a protein called Hia.<sup>5</sup> Among encapsulated strains, the primary nonpilus adhesin is a homolog of Hia called Hsf.<sup>6</sup>

Both Hia and Hsf belong to the expanding autotransporter family of proteins and are examples of so-called trimeric autotransporters, which are characterized by a trimeric architecture with three identical subunits.<sup>7–10</sup> Autotransporters are synthesized as precursor proteins with three functional domains, namely, an N-terminal signal peptide, an internal passenger domain, and a C-terminal pore-forming translocator domain.<sup>11,12</sup> The C-terminal translocator domain is embedded in the outer membrane and facilitates delivery of the internal passenger domain to the bacterial surface. In conventional autotransporters, the C-terminal translocator domain contains approximately 300 amino acids and is monomeric.<sup>9,13</sup> In contrast, in trimeric autotransporters, the translocator domain contains ~60–70 amino acids and forms trimers in the outer membrane.<sup>7,14</sup> In a recent work, Oomen *et al.* solved the crystal structure of the translocator domain of the *Neisseria meningitidis* NalP conventional autotransporter and observed a monomeric  $\beta$ -barrel pore containing 12 transmembrane  $\beta$ -strands and traversed by an N-terminal  $\alpha$ -helix.<sup>15</sup> More recently, we solved the structure of the Hia translocator domain and found that it forms a  $\beta$ -barrel with 12 transmembrane  $\beta$ -strands as well, in this case with each subunit in the trimer contributing four strands. The Hia  $\beta$ -barrel has a 1.8-nm-diameter central channel that is traversed by three N-terminal  $\alpha$ -helices, one from each subunit.<sup>16</sup>

\*Corresponding author. E-mail address:

[g.waksman@bbk.ac.uk](mailto:g.waksman@bbk.ac.uk).

Abbreviations used: MAD, multiwavelength anomalous dispersion; PEG, polyethylene glycol; ESRF, European Synchrotron Radiation Facility.

A number of models have been proposed for the translocation of autotransporter passenger domains across the outer membrane.<sup>13,15</sup> However, only two of these models are consistent with the structures of the translocator domains of NalP and Hia, namely, the “hairpin” model and the “Omp85” model.<sup>15,16</sup> According to the hairpin model, translocation of the passenger domain is initiated with the C-terminal end of the passenger domain forming a hairpin structure in the  $\beta$ -barrel and is completed as the rest of the passenger domain sequence slides through the pore. According to the Omp85 model, the pore-forming Omp85 outer-membrane protein facilitates insertion of the autotransporter translocator domain into the outer membrane and then translocates the autotransporter passenger domain through the Omp85 pore.<sup>15</sup>

Hia contains two homologous binding domains, called HiaBD1 and HiaBD2, that interact with the same host cell receptor. The nature and identity of the receptor(s) are still unknown. In the prototype Hia protein in *H. influenzae* strain 11, HiaBD1 is defined by amino acids 585–705 and binds to cultured epithelial cells with a  $K_d$  of 0.05–0.1 nM, whereas HiaBD2 is defined by amino acids 51–166 and binds to cultured epithelial cells with a  $K_d$  of 1–2 nM.<sup>17</sup> In a previous work, we determined the crystal structure of HiaBD1 and established that perturbation of subunit–subunit interactions results in disruption of the trimer and loss of adhesive activity.<sup>18</sup> The receptor-binding pocket is formed by an acidic patch that is present on all three faces of the trimeric domain, providing potential for a multivalent interaction with the host cell surface.<sup>18</sup> Although the structure of HiaBD2 is not known, a structure-based sequence alignment identified differences in residue compositions in the acidic patch that are likely to be responsible for the receptor-binding affinity lower than that for HiaBD1.<sup>18</sup> Work on the Hsf protein has identified two binding domains homologous to HiaBD1 and HiaBD2, called HsfBD1 and HsfBD2, respectively.<sup>8</sup> Based on site-directed mutagenesis, the receptor-binding pockets in HsfBD1 and HsfBD2 are formed by acidic patches that parallel the acidic patches in HiaBD1 and HiaBD2.<sup>8</sup>

In the current study, we show that the architecture of Hia and Hsf is characterized by passenger domains with multiple repeats of primarily three distinct structural domains. In addition, we compare the structures of two structural repeats with adhesive activity and one structural repeat with nonadhesive activity, shedding light on the structural determinants of Hia and Hsf adherence. We also report the structure of an extended version of the Hia translocator domain, revealing the structural transition between the translocator domain and the N-terminal passenger domain. This transition is mediated by a highly intertwined domain that is ubiquitous among trimeric autotransporters. Thus, the structures of Hia fragments presented in this report reveal a remarkable repetitive architecture in trimeric autotransporters.

## Results

### Domain arrangement in Hia-like adhesins

To gain insight into the structure of the full-length Hia protein, the Hia amino acid sequence was submitted to the daTAA server,<sup>†</sup> a server designed to analyze sequences of trimeric autotransporters.<sup>19</sup> As shown in Fig. 1, the sequence analysis predicted an impressive multirepetitive domain arrangement (Fig. 1) consisting of five domain types, with some repeated several times along the sequence. The Hia sequence contains five so-called “Trp-ring” domains (residues 117–166, 263–319, 372–421, 481–530, and 654–705, designated W1, W2, W3, W4, and W5, respectively, in Fig. 1) and three so-called “Neck” domains, including two domains with functional insertions (referred to as “IsNeck” domains; residues 53–113, 589–650, and 975–996, designated IN1, IN2, and Neck, respectively, in Fig. 1). In addition, there are two so-called “KG” domains (residues 320–372 and 743–798, designated KG1 and KG2, respectively, in Fig. 1), one so-called “GANG” domain (residues 180–263), one so-called “TTT” domain (residues 842–974), one signal peptide (residues 1–49, designated SP in Fig. 1), and one membrane anchor/translocator domain (residues 1022–1098, designated “TM” in Fig. 1). The distinctive features of the Trp-ring, Neck, IsNeck, KG, GANG, and TTT domains are well described.<sup>19</sup>

A similar analysis of the *H. influenzae* Hsf adhesin predicted a repetitive domain arrangement analogous to the arrangement in Hia, including 14 Trp-ring repeats, 4 Neck domains (with 3 IsNeck domains), 4 KG domains, 3 GANG domains, 1 TTT domain, a signal peptide, and a membrane anchor/translocator domain (Fig. 1).

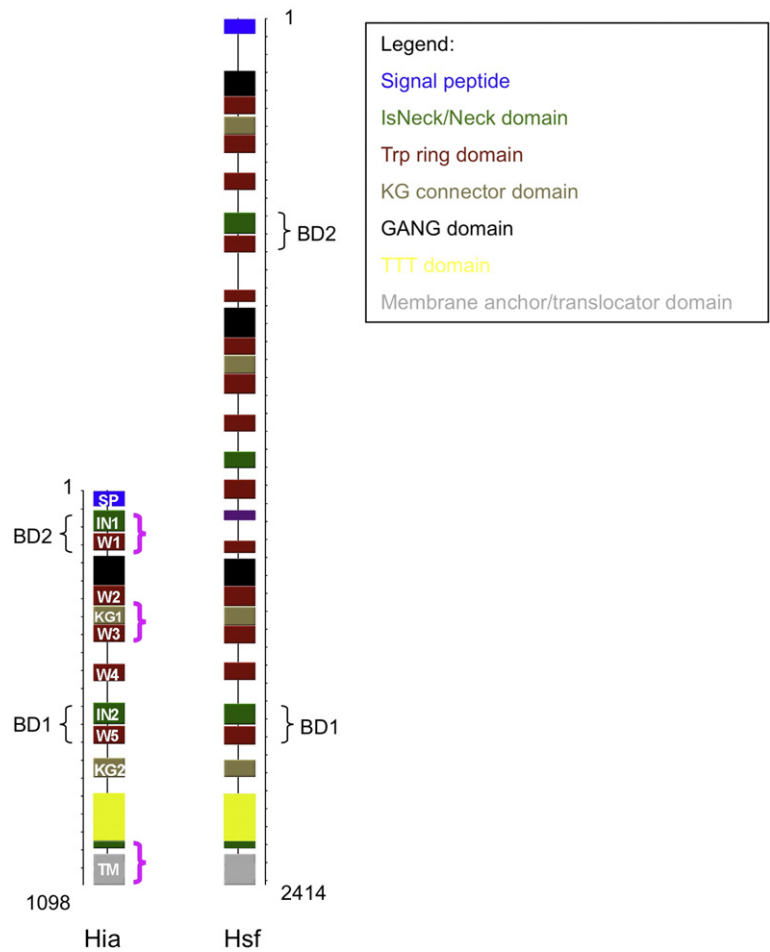
The adhesive activities in both Hia and Hsf locate to regions containing one IsNeck domain, followed by one Trp-ring domain (Fig. 1). Interestingly, not all IsNeck and Trp-ring domains are adhesive. Instead, there appears to be a requirement for immediately contiguous IsNeck and Trp-ring domains—an arrangement that is achieved in only two regions in Hia (HiaBD1, formed by IN2 and W5, and HiaBD2, formed by IN1 and W1) and two regions in Hsf (HsfBD1, formed by IN3 and W14, and HsfBD2, formed by IN1 and W4) (Fig. 1).

### Structures of Hia<sub>51–166</sub> and Hia<sub>307–422</sub>

Trp-ring domains appear to be crucial repeated modular units in Hia, both in the general architecture of the passenger domain and in the structure of the binding domains. However, the sequences of these domains are very poorly conserved, and it is not clear whether they share the same structural fold. To investigate this issue and to clarify the structural determinants of Hia-mediated binding, we deter-

<sup>†</sup> <http://toolkit.tuebingen.mpg.de/dataa>

<sup>‡</sup> <http://toolkit.tuebingen.mpg.de/dataa/browse>



**Fig. 1.** The repetitive domain arrangement of nonencapsulated *H. influenzae* Hia and encapsulated *H. influenzae* Hsf. The sequences in Hia that are determined crystallographically in this report are highlighted with purple brackets.

mined the structures of two Hia passenger domain fragments: one with adhesive activity (Hia<sub>51–166</sub>, corresponding to HiaBD2, which shares a 45% sequence identity with HiaBD1), and the other with

no adhesive function (Hia<sub>307–422</sub>, which shares an 18% sequence identity with HiaBD1).

The structure of HiaBD2 was determined to a resolution of 2.0 Å, using molecular replacement and

**Table 1.** Data collection of *H. influenzae* Hia adhesin structures

	Hia <sub>51–166</sub>	Hia <sub>307–422</sub>	Hia <sub>937–1098</sub>		
Derivative	Native	Selenomethionine	Native		
Dataset		Peak	Inflection	Remote	
Space group (molecules/asymmetric unit)	<i>I</i> 2 <sub>1</sub> 2 <sub>1</sub> (3)	<i>P</i> 6 <sub>3</sub> 22 (1)			<i>C</i> 2 (3)
Unit cell dimensions					
<i>a</i> (Å)	82.2	53.9			194.3
<i>b</i> (Å)	91.0	53.9			45.8
<i>c</i> (Å)	94.2	151.7			56.4
$\beta$ (°)	95.1				
Source/station <sup>a</sup>	ID29	ID14.4			ID14.1
Wavelength (Å)	0.9729	0.9785	0.9790	0.9750	0.9340
Resolution range (Å)	30–2.0	30–1.8	30–1.8	30–1.8	30–3.0
Observations ( <i>I</i> / $\sigma$ ( <i>I</i> )>0)	141,856	419,003	419,296	404,071	29,016
Unique reflections ( <i>I</i> / $\sigma$ ( <i>I</i> )>0)	24,079	12,834	12,852	12,885	8519
Last shell (Å)	2.11–2.00	1.86–1.80	1.86–1.80	1.86–1.80	3.11–3.00
<i>R</i> <sub>sym</sub> (%) <sup>b,c</sup>	6.5 (31.8)	9.9 (24.2)	9.8 (27.2)	10.5 (28.7)	7.9 (15.0)
Mean ( <i>I</i> / $\sigma$ ( <i>I</i> )) <sup>c</sup>	19.0 (3.7)	42.0 (11.3)	42.3 (11.3)	42.1 (11.0)	11.1 (4.1)
Completeness (%) <sup>c</sup>	99.3 (95.6)	99.6 (99.9)	99.7 (99.9)	99.7 (100.0)	84.6 (60.8)
Redundancy <sup>c</sup>	5.9 (4.1)	32.6 (28.3)	32.6 (28.2)	31.4 (26.1)	3.6 (2.4)

<sup>a</sup> Beamline designations refer to the ESRF.

<sup>b</sup>  $R_{\text{sym}} = \sum (I - \langle I \rangle) / \sum I$ .

<sup>c</sup> Values for the overall high-resolution are presented shell in parentheses.

HiaBD1 as search model (Tables 1 and 2; Supplementary Fig. 1). HiaBD2 contains two well-defined structural domains, namely, IN1 and W1. IN1 is a globular knob-like domain with a simple  $\alpha/\beta$  sandwich fold. Two short  $\beta$ -strands ( $\beta_{\text{IN1}1}$  and  $\beta_{\text{IN1}2}$ ) form a short  $\beta$ -sheet, flanked by four helices:  $\alpha_{\text{IN1}1}$ – $\alpha_{\text{IN1}3}$  on one side and  $\alpha_{\text{IN1}4}$  on the other side (Fig. 2a). In the trimer,  $\alpha_{\text{IN1}4}$  from each HiaBD2 subunit forms a three-helix bundle parallel with the axis of the trimer from which the three knob-like shapes protrude laterally (Fig. 2b). The W1 domain of HiaBD2 is an all- $\beta$  domain consisting of five long  $\beta$ -strands ( $\beta_{\text{W1}1}$ – $\beta_{\text{W1}5}$ ; Fig. 2a). In the monomer, these strands—including two  $\beta$ -hairpins ( $\beta_{\text{W1}1}$ – $\beta_{\text{W1}2}$  and  $\beta_{\text{W1}4}$ – $\beta_{\text{W1}5}$ ) and a connector strand ( $\beta_{\text{W1}3}$ )—are remarkably segregated and twisted into an N-shape (Fig. 2a). In the trimer, these strands are highly intertwined, forming a five-stranded  $\beta$ -sheet on each face of the trimer and serving as a surface against which the  $\alpha_{\text{IN1}2}$  and  $\alpha_{\text{IN1}3}$  helices of IN1 rest (Fig. 2b). The structures of HiaBD2 and HiaBD1 are very similar, aligning with a root mean square deviation (RMSD) in C $^{\alpha}$  positions of 1.1 Å between IN1 of HiaBD2 and IN2 of HiaBD1, and 1.5 Å between W1 of HiaBD2 and W5 of HiaBD1 (see superposition of the two structures in Fig. 3a).

For Hia<sub>307–422</sub>, the initial phasing efforts using either HiaBD1 or HiaBD2 as search model for molecular replacement were unsuccessful. Thus, the Ile residue at position 324 in the nonconserved region of KG1 was mutated to Met, allowing production of selenomethionine-derivatized crystals usable for phasing by the multiwavelength anomalous dispersion (MAD) method.<sup>18</sup> The crystal structure of Hia<sub>307–422</sub> was determined to a resolution of 1.8 Å.

The nonadhesive Hia<sub>307–422</sub> fragment contains three distinct domains, including the last strand of the preceding W2 domain ( $\beta_{\text{W2}5}$ ), a novel “KG1”

domain, and the W3 domain (Fig. 2c).  $\beta_{\text{W2}5}$  lies in the crystal packing interface between two Hia<sub>307–422</sub> trimers and forms an anti-parallel  $\beta$ -sheet with the adjacent  $\beta_{\text{W2}5}$  strand of a separate trimer. Although the conformation of  $\beta_{\text{W2}5}$  observed in Hia<sub>307–422</sub> might not accurately reflect the conformation it adopts in the W2 domain, it may be used to envisage the structural transition between Trp-ring domains via a KG connector domain.

The structure of the KG1 domain appears to be related to the structure of the IsNeck domain, with the secondary structures positioned similarly relative to each other. Indeed, superposition of IN2 of HiaBD1 onto KG1 of Hia<sub>307–422</sub>, obtained by aligning the two-stranded  $\beta$ -sheets of both domains (Fig. 3d), shows that the  $\alpha_{\text{KG1}1}$ ,  $\alpha_{\text{KG1}2}$ , and  $\alpha_{\text{KG1}3}$  helices are equivalent to the  $\alpha_{\text{IN2}2}$ ,  $\alpha_{\text{IN2}1}$ , and  $\alpha_{\text{IN2}4}$  helices, respectively. In both IN2 and KG1, the C-terminal helix forms an important element of the trimer interface (Fig. 2d). This structural similarity was not anticipated from sequence alignment. However, it must be noted that the topologies of the IN and KG domains (i.e., the way the secondary structures are connected) are very different (Figs. 2a and c and 3e and f).

The W3 domain in Hia<sub>307–422</sub> is similar to the W1 domain in HiaBD2 and to the W5 domains in HiaBD1, with the same number of strands and the same angle of the strands relative to the axis of the trimer (Fig. 3a). All of these strands twist and intertwine into a functional trimer, with a total buried area between the subunits well over 8000 Å<sup>2</sup> (Fig. 2b and d).

### Structure comparison between Trp-ring domains

Comparison of all of the available Trp-ring domain structures of Hia, namely, W1 (residues 117–166),

**Table 2.** Structure refinement of *H. influenzae* Hia adhesin structures

	Hia <sub>51–166</sub>	Hia <sub>307–422</sub>	Hia <sub>937–1098</sub>
Resolution range (Å)	30–2.0	30–1.8	30–3.0
R-factor (%)	19.4	16.2	23.3
R-factor (high-resolution shell) <sup>a</sup>	21.3	13.7	29.13
R <sub>free</sub> (%) <sup>b</sup>	23.6	19.8	28.5
R <sub>free</sub> (high-resolution shell)	28.1	19.7	35.2
Total number of nonhydrogen atoms	2720	1030	2617
Protein atoms	2528	786	2617
Water molecules	192	244	0
RMSD <sup>c</sup>			
Bond length (Å)	0.008	0.005	0.006
Bond angle (°)	1.079	0.881	0.942
Wilson B-factor (Å <sup>2</sup> )	30.3	15.8	56.4
Average B-factor protein atoms (Å <sup>2</sup> )	44.3	17.8	48.8
Average B-factor solvent atoms (Å <sup>2</sup> )	44.4	32.6	
Ramachandran statistics <sup>d</sup>			
Most favored regions (%)	90.9	93.6	82.5
Additionally allowed regions (%)	9.1	6.4	17.1
Generously allowed regions (%)	0	0	0.3
Disallowed regions (%)	0	0	0

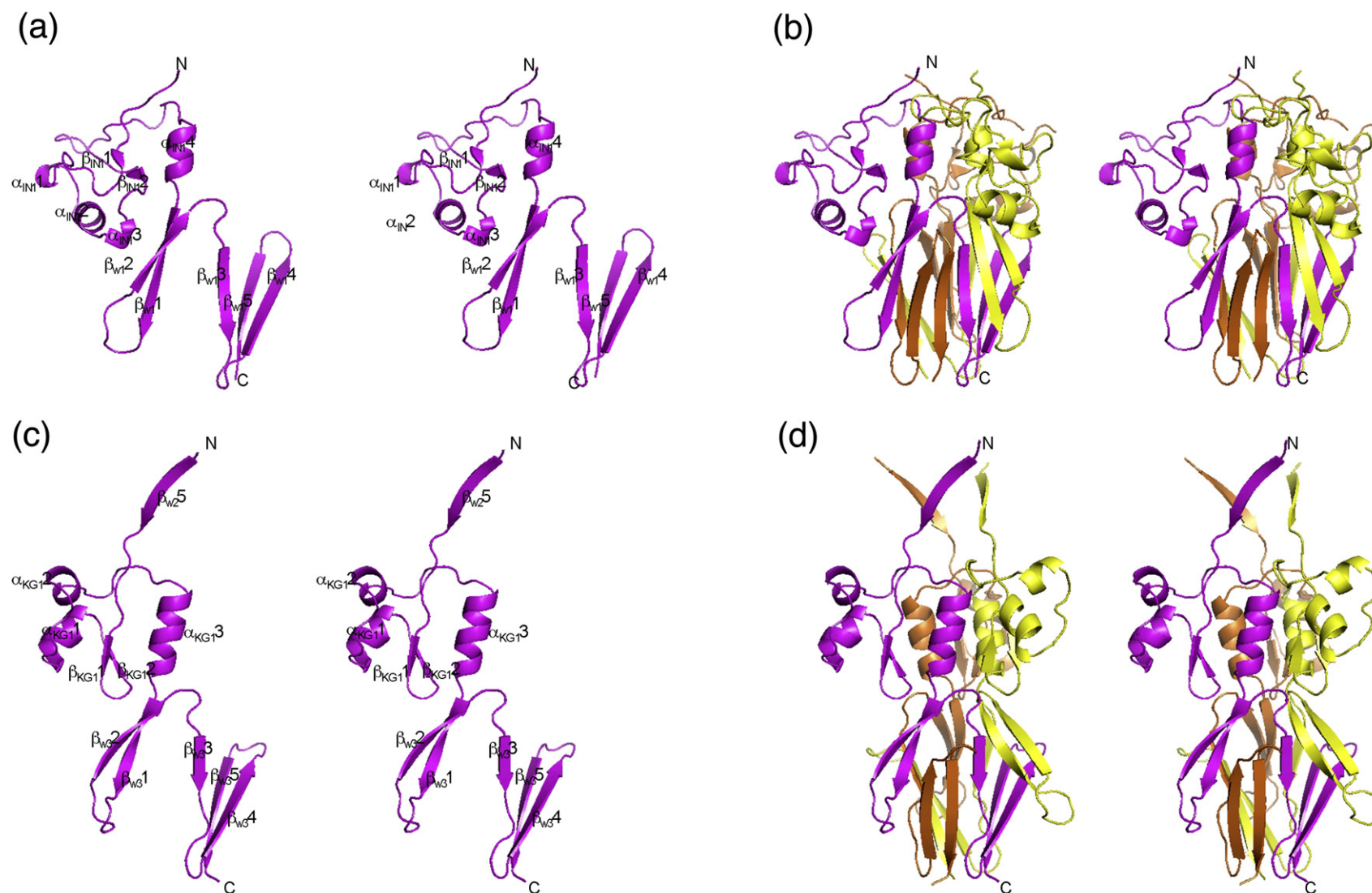
<sup>a</sup> High-resolution shell (see the last shell in Table 1).

<sup>b</sup> R<sub>free</sub> calculated using 5% or 10% of the total reflections omitted from refinement for Hia<sub>51–166</sub>, Hia<sub>307–422</sub>, and Hia<sub>937–1098</sub>.

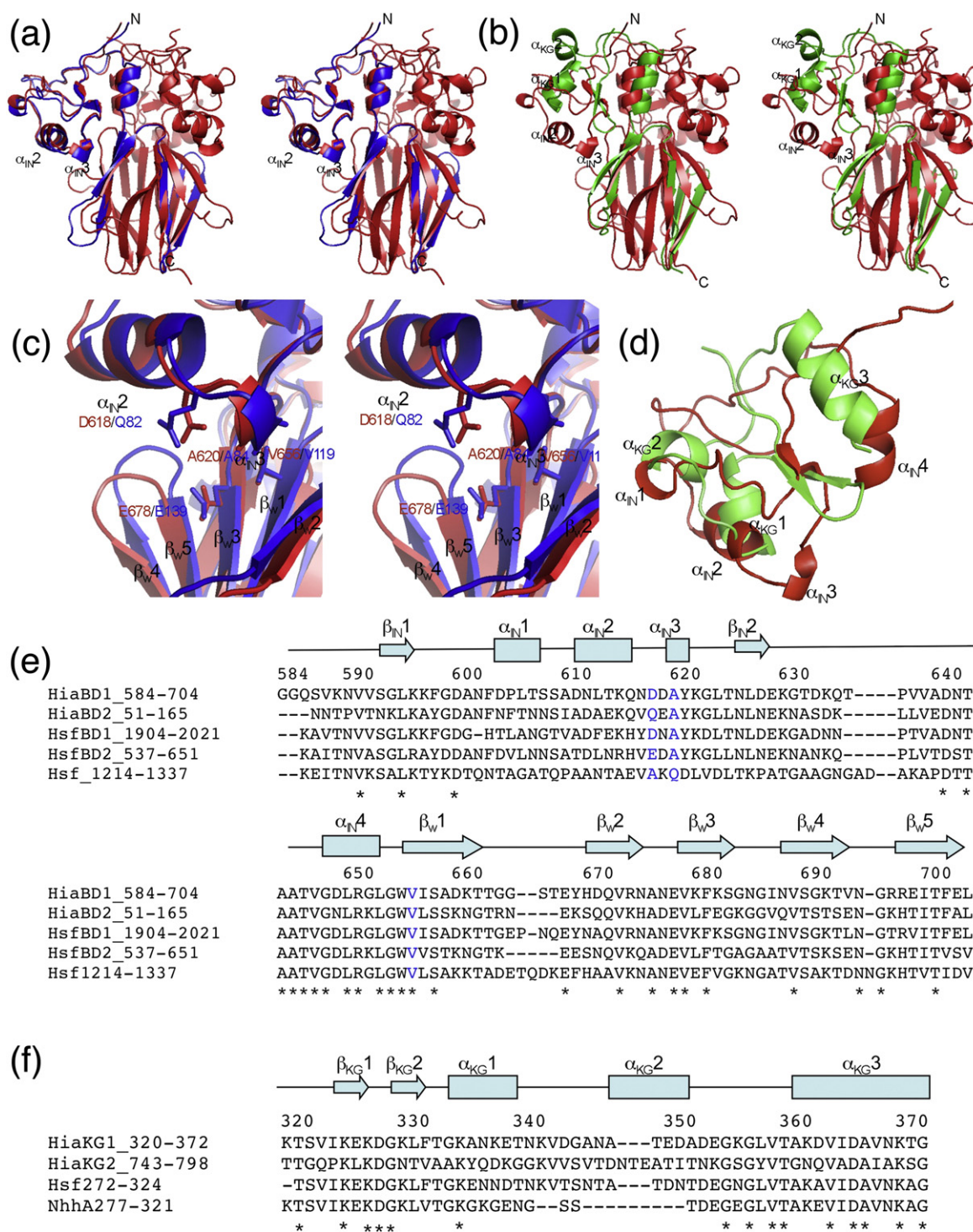
<sup>c</sup> RMSDs from ideal bond lengths and angles between bonded atoms.<sup>37</sup>

<sup>d</sup> Ramachandran statistics calculated using PROCHECK.<sup>38</sup>





**Fig. 2.** Structures of the Hia<sub>51-166</sub> and Hia<sub>307-422</sub> monomers and trimers. (a) Stereo ribbon diagram of the Hia<sub>51-166</sub> monomer. Secondary structures, including helices and strands, are labeled. (b) Stereo ribbon diagram of the Hia<sub>51-166</sub> trimer. The three subunits are shown in magenta, yellow, and orange. (c) Stereo ribbon diagram of the Hia<sub>307-422</sub> monomer. Secondary structures, including helices and strands, are labeled. (d) Stereo ribbon diagram of the Hia<sub>307-422</sub> trimer. The three subunits are shown in the same color scheme as mentioned above.

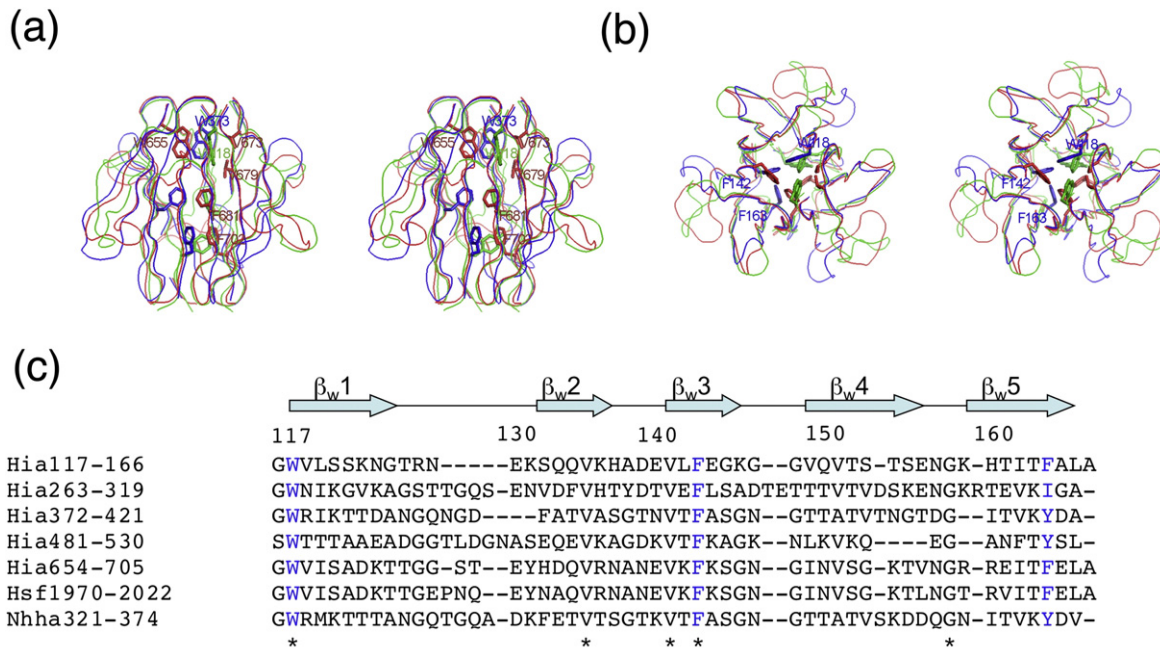


**Fig. 3.** Structure comparison between adhesive domains HiaBD1 and HiaBD2 and the nonadhesive domain Hia<sub>307-422</sub>. (a) Stereo ribbon diagram of the superimposed HiaBD1 (red) and HiaBD2 (blue) domains. (b) Stereo ribbon diagram of the superimposed HiaBD1 (red) and Hia<sub>307-422</sub> (green) domains. (c) Detailed molecular feature of the groove formed by the base of the IN domain ( $\alpha_{IN2}$  and  $\alpha_{IN3}$ ) and the top of the adjacent W domain ( $\beta_{W1}$  and  $\beta_{W3}$ ). Mutations of residues D618 (in the loop between  $\alpha_{IN2}$  and  $\alpha_{IN3}$ ), A620 (in  $\alpha_{IN3}$ ), and V656 (in  $\beta_{W1}$ ) all abolished the adhesion activities of HiaBD1 *in vivo* and *in vitro*. (d) Ribbon diagram of the superimposed IN2 (red) and KG1 (green) domains. The helices, including  $\alpha_{KG1}$ ,  $\alpha_{KG2}$ , and  $\alpha_{KG3}$  in KG domain and  $\alpha_{IN2}$ ,  $\alpha_{IN1}$ , and  $\alpha_{IN4}$  in IN domain, are labeled respectively. (e) Sequence alignment of the IN and W adhesive-like domains among Hia/Hsf/NhhA. (\*) Invariant residues. The three functional important residues are highlighted in blue. (f) Sequence alignment of the KG domains between Hia/Hsf/NhhA.

W3 (residues 372-421), and W5 (residues 654-705), reveals only minor differences (Fig. 4). The RMSDs of the main-chain atoms of these Trp-ring domains

are 1.6 Å between W1 and W3, 1.7 Å between W3 and W5, and 1.5 Å between W1 and W5. The major variations among these structures lie in the





**Fig. 4.** Structure comparison between Trp-ring domains. (a) Stereo ribbon diagram of the superimposed W1 (blue), W3 (green), and W5 (red) domains viewed from the side. (b) Stereo ribbon diagram of the superimposed Trp-ring domains viewed from the top. (c) Sequence alignment of the Trp-ring domains between Hia/Hsf/NhhA. In (a) and (b), the conserved residues at the trimeric interface are shown in stick representation. In (c), these residues are shown in blue.

$\beta_{W1}$ - $\beta_{W2}$ ,  $\beta_{W3}$ - $\beta_{W4}$ , and  $\beta_{W4}$ - $\beta_{W5}$  loops away from the axial hydrophobic cavity of the trimer (Fig. 4). Overall, the most highly conserved residues map at the trimeric interface (Fig. 4b and c). Two invariant aromatic residues in the  $\beta_{W1}$  and  $\beta_{W3}$  strands of the W1, W3, and W5 Trp-ring domains (W118/W373/W655 and F142/F398/F681, respectively) are found in immediate trimeric contact in the upper part of the Trp-ring domain. Another two invariant Val residues in the  $\beta_{W2}$  and  $\beta_{W3}$  strands (V134/V390/V673 and V140/V396/V679) also map to this region and could play a supportive role underneath the invariant Trp residues (Fig. 4c). Finally, a third less conserved aromatic patch in the  $\beta_{W5}$  strand is clearly visible in the lower part of the W1, W3, and W5 domains (F163/Y419/F702) (Fig. 4). The only W domain in which this aromatic ring is not seen is W2, where the trimeric interface is mediated by Ile residues (Fig. 4c).

#### Structure comparison between HiaBD1 and HiaBD2 adhesive domains and the Hia<sub>307-422</sub> nonadhesive domain

In our previous report of the HiaBD1 structure, we identified the surfaces responsible for binding activity (Fig. 3c). This region is located in a groove formed by the base of the IN domain ( $\alpha_{IN2}$  and  $\alpha_{IN3}$ ) and the top of the W domain ( $\beta_{W1}$  and  $\beta_{W3}$ ). Notably, individual mutations of residues D618 (in the loop between  $\alpha_{IN2}$  and  $\alpha_{IN3}$ ), A620 (in  $\alpha_{IN3}$ ), and V656 (in  $\beta_{W1}$ ) abolished the adhesive activity of HiaBD1 (Fig. 3c and e).<sup>18</sup>

While HiaBD1 has structural similarities with Hia<sub>307-422</sub>,  $\alpha_{IN3}$  in HiaBD1 does not have a corresponding region in KG1 in Hia<sub>307-422</sub>, and  $\alpha_{IN2}$  in HiaBD1 is oriented differently from  $\alpha_{KG1}$  in Hia<sub>307-422</sub> (Fig. 3b and d). Thus, two important structural determinants of binding in the IN domain are either absent or misoriented in the KG domain, explaining why Hia<sub>307-422</sub> is nonadhesive.

In an earlier work, we found that HiaBD1 and HiaBD2 recognize the same host cell receptor but with different binding affinities.<sup>17</sup> However, due to the lack of structural information on HiaBD2, it was not clear which residues account for this difference in adhesive activity. The examination of the binding groove described above (Fig. 3c and e) reveals only one notable difference among functionally important residues between HiaBD1 and HiaBD2 in this region, namely, D618 in HiaBD1 corresponds to a glutamine residue in HiaBD2, resulting in a change in charge and a longer side chain. The corresponding residues in HsfBD1 and HsfBD2 are aspartic acid and glutamic acid, respectively. HsfBD1 and HsfBD2 interact with the same host cell receptor recognized by HiaBD1 and HiaBD2, with HsfBD1 exhibiting a binding affinity higher than that of HsfBD2,<sup>8</sup> suggesting that the extended side chain of glutamic acid may result in steric hindrance (Fig. 3e). Interestingly, Hsf<sub>1214-1337</sub> is a region that contains an IsNeck domain and a Trp-ring domain—an arrangement that appears to determine adhesive function. However, in Hsf<sub>1214-1337</sub>, the residues corresponding to D618 and A620 in HiaBD1 are Ala and Gln (Fig. 3e). Consistent with the information on the binding pockets in HiaBD1, HiaBD2,

HsfBD1, and HsfBD2, Hsf<sub>1214–1337</sub> lacks adhesive activity.<sup>8</sup>

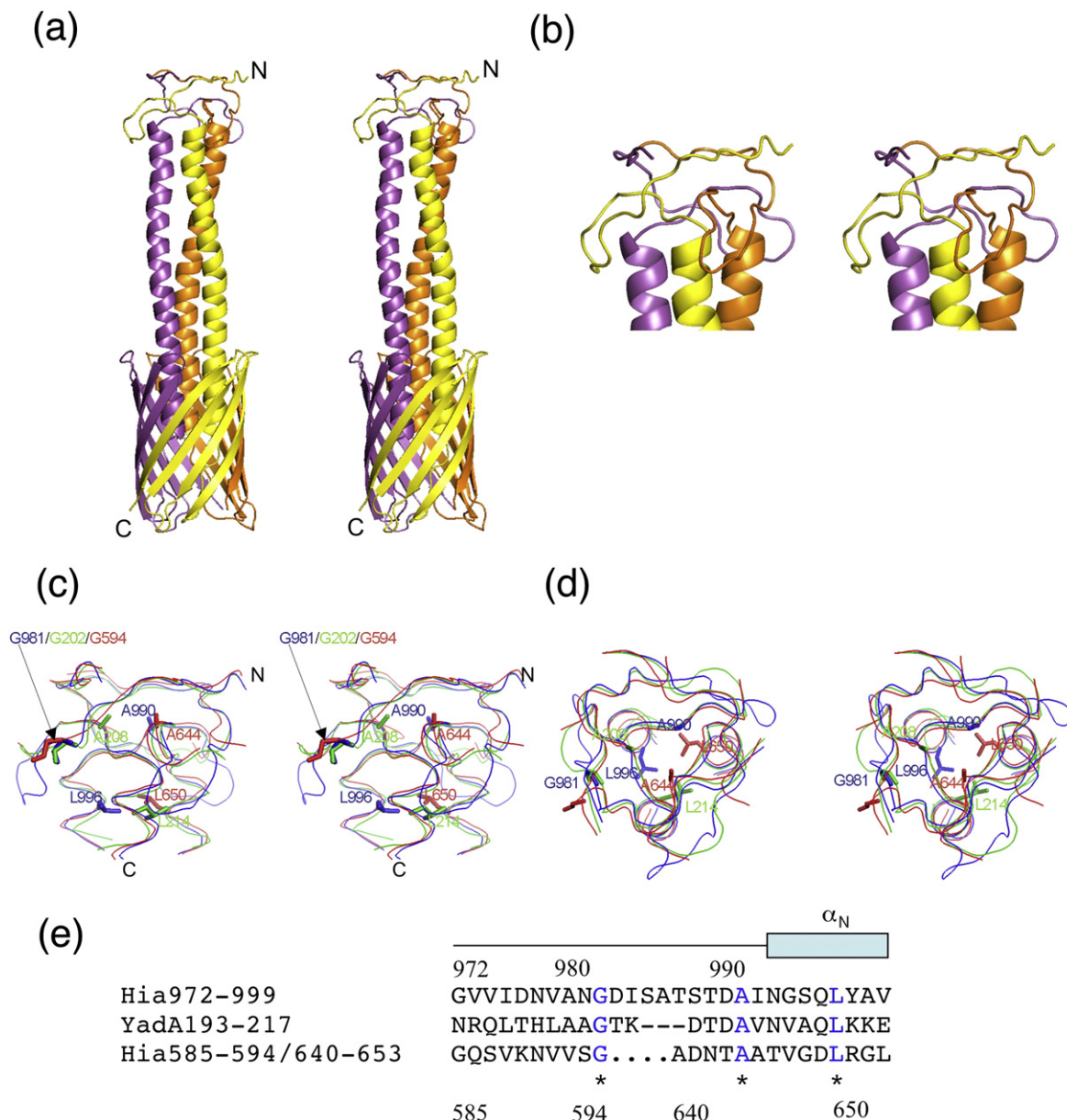
### Transmembrane structure of Hia<sub>973–1098</sub>

Although the coding sequence used for the expression and purification of the Hia translocator domain contained residues 937–1098, only the last 126 residues were visible in the electron density map (Hia<sub>973–1098</sub>). The structure was determined to a resolution of 3.0 Å.

Compared to the previously published Hia<sub>998–1098</sub> structure, the structure of Hia<sub>973–1098</sub> reveals the full extent of the helices passing through the β-barrel

and an extra highly intertwined Neck domain (Hia<sub>Neck</sub>) sitting on top (Fig. 5a and b). Neck domains are short connector domains (~20–30 residues; Fig. 5e) between α-helical coiled coils and β-regions<sup>18,20</sup> that are often interrupted by the insertion of sequences to form IsNeck domains.<sup>19,21</sup>

The Hia<sub>973–1098</sub> trimer is formed through juxtaposition of the transmembrane β-sheets (residues 1043–1098), stacking of the long α-helices (residues 992 and 1036), and interactions between the long loops of the Neck domain along a 3-fold axis running through the center of the trimeric structure (residues 972 and 991) (Fig. 5a). The transmembrane parts of Hia<sub>973–1098</sub> are virtually the same as those in



**Fig. 5.** Transmembrane structure of Hia<sub>973–1098</sub>. (a) Stereo ribbon diagram of the Hia<sub>973–1098</sub> trimer. The three subunits are shown in magenta, yellow, and orange. (b) Stereo ribbon diagram of the Hia<sub>Neck</sub> domain. (c) Stereo ribbon diagram of the superimposed domains Hia<sub>Neck</sub> (blue), Neck in IN2 (red), and YadA<sub>Neck</sub> (green) viewed from the side. (d) Stereo ribbon diagram of the same superimposed Neck domains viewed from the top. (e) Sequence alignment of the Neck domains between Hia<sub>Neck</sub>, Neck in IN2, and YadA<sub>Neck</sub>.

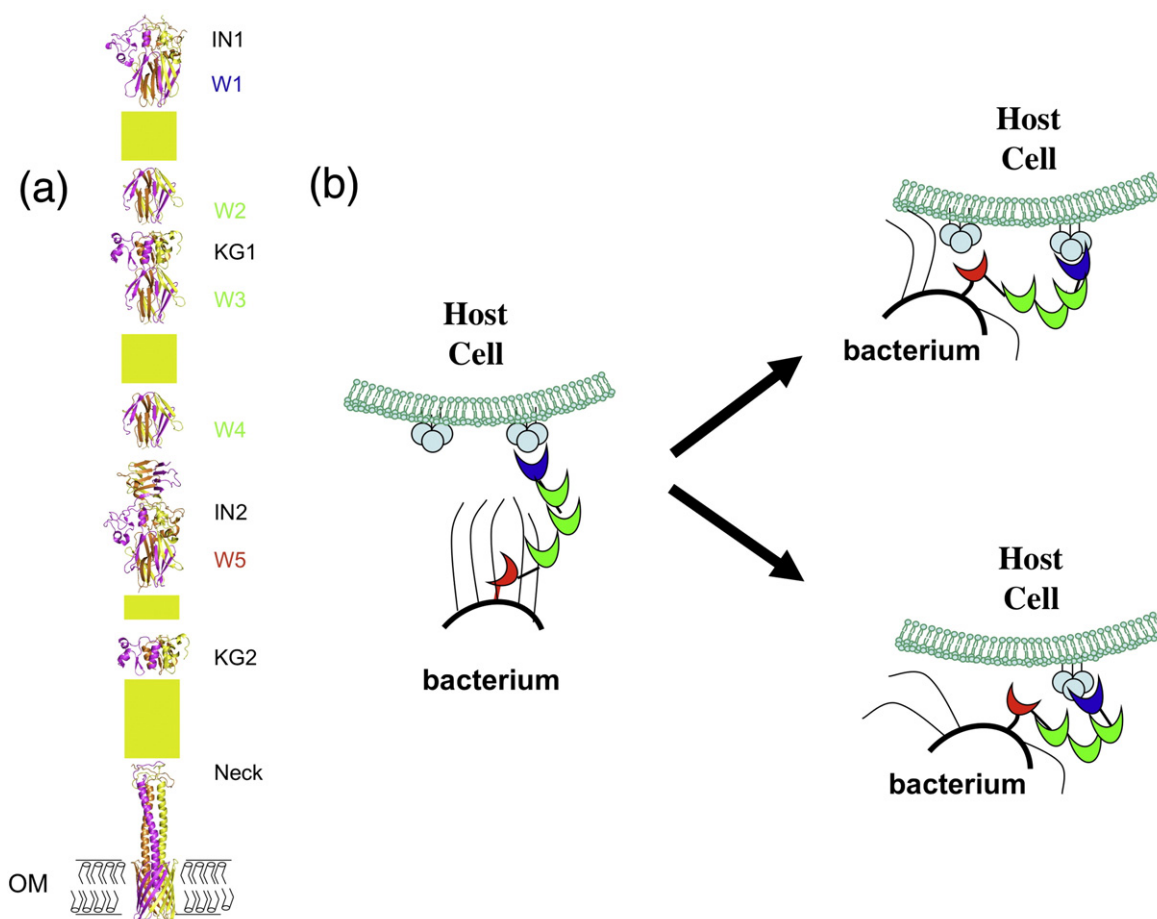


Hia<sub>998–1098</sub>. The three  $\alpha$ -helices protrude from the 12-stranded  $\beta$ -barrel and form a three-helix coiled coil. The last residues in all three helices introduce a 90° kink, followed by an extended hairpin loop. There are only a few intramolecular contacts within Hia<sub>Neck</sub>; thus, the Neck domain appears to hold together via its neighboring subunits through side-chain, van der Waals, and main-chain hydrogen-bonding interactions (Fig. 5c and d). Comparison of Hia<sub>Neck</sub> with the necks of Hia<sub>IN2</sub> and YadA (YadA<sub>Neck</sub>) shows that all three are very similar. The RMSDs between the main-chain atoms of these Neck-like domains are 1.5 Å between Hia<sub>Neck</sub> and Hia<sub>IN2</sub>, 2.5 Å between Hia<sub>IN2</sub> and YadA<sub>Neck</sub>, and 1.0 Å between Hia<sub>Neck</sub> and YadA<sub>Neck</sub>. Structure-based sequence alignment reveals three invariant residues in Hia<sub>Neck</sub>, YadA<sub>Neck</sub>, and Hia<sub>IN2</sub>, including G981/G202/G594, A990/A208/A644, and L996/L214/L650 (Fig. 5c–e). G981/G202/G594 lies at the turn of the hairpin-like loops away from the axis of the overall trimeric structure and does not contribute any van der Waals or hydrogen-bonding interactions.

Interestingly, these glycines are also the sites of major sequence insertions in the IsNeck domains. A990/A208/A644 and L996/L214/L650 always map axially in the trimeric packing interface, playing important roles in holding the trimeric Neck domain together.

## Discussion

Following determination of the structures of HiaBD1<sup>18</sup> and the Hia translocator domain,<sup>16</sup> it became clear that trimeric autotransporters share specific structural features and also have distinct structural differences compared with conventional autotransporters. Among the similarities is the translocator domain, which consists of a conserved 12-stranded  $\beta$  barrel with a central channel traversed by one or three  $\alpha$  helices. Among the distinguishing features is the highly singular fold of the adhesive domains of trimeric autotransporters, where the fold of each of the subunits can be understood only in the context of the trimer due to the highly intertwined nature of the trimeric interface.



**Fig. 6.** Modular architecture of Hia and putative two-step adhesive mechanism. (a) Modular architecture of Hia adhesin. While the structures of IN1, W1, KG1, W3, IN2, W5, Neck, and transmembrane anchor/translocator domains are determined crystallographically, W2/W4 and KG2 are modeled based on W1/W3/W5 and KG1, respectively. Yellow rectangles represent the Hia sequences that are yet to be structurally characterized, including the predicted N-terminal GANG and C-terminal TTT domains. (b) A putative two-step adhesive mechanism utilized by Hia/Hsf-like adhesin to form an intimate association between the bacterium and the host cell. The adhesive Trp-ring domains with different binding capacities are shown in blue and red. The nonadhesive Trp-ring domains are shown in green.

In this study, we identified novel distinguishing features of Hia. In particular, the Hia passenger domain consists of repeats of structurally similar domains, including five structurally identical Trp-ring domains strung together by two IsNeck domains and two KG domains (Fig. 6). Furthermore, we found that the IsNeck domains and the KG domains share a similar three-dimensional organization, although they differ topologically. At the C-terminus, the multidomain structure of the passenger domain is connected to the Hia membrane anchor/translocator domain via an  $\alpha$ -helical coiled-coil-like structure preceded by a Neck connector domain.

While the Neck domains appear to be highly conserved in trimeric autotransporters (particularly at the junction of the passenger and translocator domains), Trp-ring and IsNeck domains appear to be present in only a few of these proteins, and KG domains are even less common (Supplementary Fig. 2). Interestingly, Trp-ring and IsNeck domains appear to be linked evolutionarily, as there are no known trimeric autotransporters exhibiting one without the other. This feature may be related to the observation that both domains must be juxtaposed to display adhesive activity (although this arrangement is not sufficient for adhesive activity, as highlighted by Hsf<sub>1214–1337</sub>).

The implications of such a modular architecture are twofold. First, while conventional autotransporters adopt a relatively simple  $\beta$ -helical arrangement, the folding of trimeric autotransporters is much more complex, not only requiring simultaneous processing of all three subunits but also following a complex folding pathway involving multiple structural domains of different types. Second, the modular architecture of trimeric autotransporters may provide a versatile structural template that can be utilized to create new isoforms by gene shuffling, duplication, or recombination. Hsf may be a good illustration of this concept. The Hsf adhesin includes 14 Trp-ring repeats chained together by three IsNeck domains, four KG domains, and three GANG domains. However, the domain organization at the C-terminus of the Hsf passenger domain is remarkably similar to the organization of Hia, while the N-terminus of Hsf resembles NhhA, an *N. meningitidis* trimeric autotransporter. Thus, Hsf could have arisen from duplication of the Hia gene and capture of an NhhA-like gene (Supplementary Fig. 2). A similar chimeric architecture has been proposed for the head domain of the trimeric autotransporter BadA from *Bartonella henselae*.<sup>22</sup>

Both Hia and Hsf contain two binding domains (HiaBD1/HsfBD1 and HiaBD2/HsfBD2) that are located at opposite ends of the passenger domain and have different binding affinities. Interestingly, in both cases, the binding domain at the C-terminal end of the protein is associated with greater affinity. The work presented here suggests that the difference in affinity is most likely explained by changes in the composition of the binding pocket

identified in HiaBD1.<sup>18</sup> Thus, due to their trimeric structure, Hia and Hsf display a multivalent binding site with six subsites (three HiaBD1/HsfBD1 subsites and three HiaBD2/HsfBD2 subsites). As outer-membrane adhesins, the functions of Hia and Hsf are to make contact with host cell receptors and to facilitate interaction between the bacterium and the host. To achieve this goal, the adhesin must be long enough to extend beyond the lipopolysaccharide layer in the outer membrane of the bacterium and to penetrate the glycan layer surrounding the host cell. In addition, the adhesin must be flexible in order to form an intimate interaction with the host and to surmount mechanical forces in the respiratory tract. Based on the available structures of Hia<sub>51–166</sub> (~53 Å long), Hia<sub>307–422</sub> (~75 Å long), Hia<sub>541–705</sub> (~80 Å long), and Hia<sub>973–1098</sub> (~92 Å long), the estimated length of the Hia passenger domain is at least 600 Å. The estimated length of Hsf is at least 1200 Å. Because Hsf is present in encapsulated strains of *H. influenzae*, this adhesin may indeed need to be longer to reach beyond the capsule.

The repetitive domain architecture and the two groups of binding pockets in Hia and Hsf suggest a two-step mechanism for host recognition and attachment (Fig. 6b). We propose that the distal binding domain (HiaBD2/HsfBD2) initiates the initial interaction with the host cell surface. Subsequently, the repetitive domain architecture between the two adhesive domains, reminiscent of the repetitive Ig-fold architectures of intimin or P and type 1 pili,<sup>23–25</sup> provides the flexibility required to allow the HiaBD1/HsfBD1 domain to compete with HiaBD2/HsfBD2 for the very same receptor or to interact with a second receptor on the host cell surface. Via this two-step mechanism, the bacterium is pulled closer towards the host cell and forms an intimate association that is strong enough to surmount the mechanical defensive force in the host.

The vast size and modular architecture of Hia and Hsf must have an impact on their mechanism of secretion through the outer membrane. To date, two secretion models appear to be consistent with the known structural biology of trimeric autotransporters: the so-called “hairpin” and “Omp85” models. Recent advances in the structural biology of the DegP chaperone suggest a variation in the Omp85 model in which DegP forms a hydrophobic cage vast enough to accommodate folded outer-membrane proteins, captures these proteins in the periplasm, and then delivers them to the Omp85/YaeT complex for insertion into the outer membrane.<sup>26</sup> Folding of membrane proteins within the periplasm has been suggested by studies of PhoE,<sup>27</sup> EspP,<sup>28</sup> and Hia.<sup>16</sup> One objection to this mechanism being applied to Hia and Hsf is that these proteins are too long to be accommodated within the DegP cage. However, it is possible that the DegP cage surrounds only the translocator domain and carries out the periplasmic shuttling function, with the folded hydrophilic passenger domain protruding from the cage.

## Materials and Methods

### Bacterial strains, plasmids, and culture conditions

Laboratory strains employed in these studies include *Escherichia coli* B834(DE3) (methionine auxotroph; Novagen), *E. coli* DH5 $\alpha$  (Life Technologies), and *E. coli* BL21 (DE3). *E. coli* strains were grown in Luria-Bertani (LB) agar or in LB and stored at  $-80^{\circ}\text{C}$  in LB with 50% glycerol. Antibiotic concentrations used to select for plasmids included  $100\text{ }\mu\text{g ml}^{-1}$  ampicillin and  $12.5\text{ }\mu\text{g ml}^{-1}$  tetracycline.

### Construction of plasmids used in this study

To generate constructs for crystallography experiments, *H. influenzae* genomic DNA was used as template to amplify the coding sequence for Hia residues 51–166, 307–422, and 937–1098, with either a BamHI site and an NdeI site or two BsaI sites at the ends of the forward and reverse primers, respectively. The primers were as follows:

5'-GACGACATATGAACAATACTCTTA TTACGAATAAG-3' (forward primer; Hia<sub>51–166</sub>)  
 5'-GACGAGGATCCTTACGCTAAAGCAAAGGTAATGG-TGTG-3' (reverse primer; Hia<sub>51–166</sub>)  
 5'-GACGACATATGAAAGAAAACGGTAAGAGAACGAA-3' (forward primer; Hia<sub>307–422</sub>)  
 5'-GACGAGGATCCTTATTTTCGCATCATACTTAACGGTA-AT-3' (reverse primer; Hia<sub>307–422</sub>)  
 5'-ATGGTAGGTCTCACTCCGACGGTACGGCTGATATGAC-CA-3' (forward primer; Hia<sub>937–1098</sub>)  
 5'-ATGGTAGGTCTCATATCACTGGTAACCAACAC-CTGCTG-3' (reverse primer; Hia<sub>937–1098</sub>).

The resulting PCR fragments encoding Hia<sub>51–166</sub> and Hia<sub>307–422</sub> were digested with BamHI and NdeI and ligated into BamHI/NdeI-digested pET15b (Novogen), generating pET15b-Hia<sub>51–166</sub> and pET15b-Hia<sub>307–422</sub>. In order to obtain experimental phases for structural determination, a point mutation in Hia<sub>307–422</sub> converting the Ile at position 324 into Met was generated using primers 5'-ACTTCTGTTATGAAA-GAAAAAGAC-3' (forward) and 5'-GTCTTTTCTTCA-TAACAGAGT-3' (reverse), and the QuickChange™ site-directed mutagenesis kit (Stratagene). (The base mutation responsible for the amino acid is underlined.)

To generate a plasmid to overexpress Hia<sub>937–1098</sub> in the outer membrane of *E. coli*, the OmpA signal peptide encoded in pASK-IBA12 (IBA) was used. The PCR fragment of Hia<sub>937–1098</sub> was digested with BsaI and ligated into BsaI-digested pASK-IBA12, generating pASK-Hia<sub>937–1098</sub>.

### Expression, purification, and crystallization

*E. coli* BL21(DE3)/pET15b-Hia<sub>51–166</sub> and methionine auxotroph *E. coli* B834(DE3)/pET15b-Hia<sub>307–422</sub>I324M were grown at  $37^{\circ}\text{C}$  to an OD<sub>600</sub> of 0.5 and then induced with 1 mM IPTG for 4 h. Selenomethionine-derivatized Hia<sub>307–422</sub> was expressed in cells grown in SelenoMet Medium Base (AthenaES) supplemented with SelenoMet Nutrient Mix (AthenaES) and  $40\text{ mg l}^{-1}$  seleno-L-methionine (Acros). Following induction, bacteria were centrifuged at  $4000g$  for 20 min, and cell pellets were resuspended in 20 mM sodium phosphate (pH 7.4),

500 mM NaCl, and 40 mM imidazole. Bacteria were disrupted with a high-pressure homogenizer (Glen Creston Ltd.). The clear lysate ( $30,000g$  for 30 min) was loaded onto a HisTrap column (GE Healthcare) and eluted with 20 mM sodium phosphate (pH 7.4), 500 mM NaCl, and 500 mM imidazole. The eluate was pooled and incubated with thrombin protease (Novagen) at room temperature for 10 h and then dialyzed against 20 mM Tris (pH 8.0) and 20 mM NaCl at  $4^{\circ}\text{C}$  for 10 h. Subsequently, the protein sample was loaded back onto a HisTrap column, and the flow-through was loaded onto a Q-Sepharose column. A linear salt gradient (pH 8.0; 0–1 M) was used for elution. The peak fractions ( $\sim 0.2\text{ M}$  for Hia<sub>51–166</sub> and  $\sim 0.3\text{ M}$  for Hia<sub>307–422</sub>) were pooled, and purity was further polished by an S100 gel-filtration column. The final purified protein sample was estimated to be  $\sim 95\%$  pure by Coomassie-blue-stained SDS-PAGE. The eluate of the S100 was subsequently concentrated to  $\sim 47\text{ mg/ml}$  for Hia<sub>51–166</sub> and to  $\sim 62\text{ mg/ml}$  for Hia<sub>307–422</sub> using a Centricon concentrator with a 10-kDa cutoff (Amicon).

To purify Hia<sub>937–1098</sub>, *E. coli* BL21(DE3)/pASK-Hia<sub>937–1098</sub> was grown at  $37^{\circ}\text{C}$  and induced with  $0.2\text{ mg l}^{-1}$  anhydrotetracycline (IBA) for 14 h. Following induction, bacteria were centrifuged at  $4000g$  for 20 min, and cell pellets were resuspended in 20 mM Tris (pH 8.0) and 150 mM NaCl. Bacteria were disrupted with a high-pressure homogenizer (Glen Creston Ltd.). The membrane proteins were recovered by centrifugation at  $30,000g$  for 45 min and solubilized with 5% (wt/vol) Elugent (Calbiochem), 20 mM Tris (pH 8.0), and 150 mM NaCl. The clear lysate ( $30,000g$  for 45 min) was loaded onto a Strep-Tactin column and eluted with 2.5 mM desthiobiotin, 100 mM Tris (pH 8.0), 150 mM NaCl, 10 mM ethylenediaminetetraacetic acid, and 0.5% (wt/vol) Elugent. The eluate was then incubated with thrombin protease for  $>10\text{ h}$  at room temperature before it was dialyzed against 20 mM Tris (pH 8.0), 20 mM NaCl, and 0.5% (wt/vol) Elugent at  $4^{\circ}\text{C}$ . The resulting protein sample was then reloaded onto a fresh Strep-Tactin column to eliminate the cleaved Strep tag. The flow-through of the Strep-Tactin column was pooled and loaded onto a Q-Sepharose column. A linear salt gradient (pH 8.0; 0–1 M) was used for elution. The peak fractions ( $\sim 0.2\text{ M}$ ) were pooled and dialyzed against 20 mM 4-morpholineethanesulfonic acid (pH 5.2), 20 mM NaCl, and 0.5% (wt/vol) Elugent, and then loaded onto an SP-Sepharose column. A linear salt gradient (pH 5.2; 0–1 M) was used for elution. The peak fraction ( $\sim 0.5\text{ M}$ ) was  $\sim 95\%$  pure, as indicated by Coomassie-blue-stained SDS-PAGE. Finally, SP-Sepharose was again used to exchange Elugent with 0.6% (vol/vol) *n*-octyltetraosyethylene (Bachem). The eluate of SP-Sepharose was subsequently concentrated to  $\sim 29\text{ mg/ml}$  using a Centricon concentrator with a 10-kDa cutoff (Amicon) and dialyzed against 20 mM 4-morpholineethanesulfonic acid (pH 5.2), 100 mM NaCl, and 0.6% (vol/vol) *n*-octyltetraosyethylene at  $4^{\circ}\text{C}$  for 14 h.

Hia<sub>51–166</sub> native crystals, Hia<sub>937–1098</sub> native crystals, and Hia<sub>307–422</sub> selenomethionine-derivatized crystals were grown at  $20^{\circ}\text{C}$  using the hanging-drop vapor-diffusion method. The reservoir solutions are 100 mM Hepes (pH 7.5), 25% polyethylene glycol (PEG) 10,000 (Hia<sub>51–166</sub>), and 100 mM Tris (pH 8.5); 30% PEG 1000 (Hia<sub>307–422</sub>) and 100 mM Hepes (pH 7.0); and 16% PEG 2000 MME (Hia<sub>937–1098</sub>). Crystals of Hia<sub>51–166</sub> diffracted to  $2.0\text{ }\text{\AA}$  resolution and belonged to space group  $I2_12_1$ , with cell dimensions of  $a = 82.2\text{ }\text{\AA}$ ,  $b = 91.0\text{ }\text{\AA}$ , and  $c = 94.2\text{ }\text{\AA}$ , and with three molecules in the asymmetric unit. Crystals of selenomethionine-derivatized Hia<sub>307–422</sub> diffracted to



1.8 Å and belonged to space group  $P6_322$ , with cell dimensions of  $a=53.9$  Å,  $b=53.9$  Å, and  $c=151.7$  Å, and with one molecule in the asymmetric unit. Crystals of Hia<sub>937–1098</sub> diffracted to 3.0 Å resolution and belonged to space group  $C2$ , with cell dimensions of  $a=194.3$  Å,  $b=45.8$  Å,  $c=56.4$  Å, and  $\beta=95.1$ , and with three molecules in the asymmetric unit.

### Data collection and phasing

Diffraction data for Hia<sub>51–166</sub> and Hia<sub>937–1098</sub> native crystals were recorded on beamlines ID29 and ID14.1 at the European Synchrotron Radiation Facility (ESRF; Grenoble, France), respectively. MAD diffraction data of selenomethionine-derivatized Hia<sub>307–422</sub>I324M were recorded at selenium edge (0.9785 Å), inflection (0.9790), and remote (0.9750) on beamline ID14.4 at the ESRF. Hia<sub>307–422</sub>I324M MAD data were integrated and scaled using DENZO and SCALEPACK.<sup>29</sup> MOSFLM/SCALA<sup>30</sup> was used for Hia<sub>51–166</sub> and Hia<sub>937–1098</sub> diffraction data. The statistics of data collection are reported in Table 1.

The MAD method was used to phase Hia<sub>307–422</sub>I324M. One Se position was determined by anomalous-difference Patterson analysis of data with resolutions of between 30 and 3.0 Å, using the program SHELX.<sup>31</sup> The parameters of the heavy-atom sites were refined using SHARP (Global Phasing).<sup>32</sup> An interpretable map was obtained after solvent flattening (DM).<sup>30</sup> The extended phases (1.8 Å) allowed automatic tracing of the backbone (Hia residues 320–422; using the program ARP/wARP<sup>33</sup>), producing a  $\sigma_A$ -weighted  $2F_o - F_c$  map of excellent quality into which side chains were built manually using the program COOT<sup>30</sup> (Supplementary Fig. 1). Hia<sub>51–166</sub> and Hia<sub>937–1098</sub> were phased by molecular replacement (PHASER)<sup>30</sup> using the published structures of Hia<sub>585–705</sub> and Hia<sub>998–1098</sub>, respectively (Protein Data Bank codes 1S7M and 2GR7, respectively) as search models. ARP/wARP and COOT were used as mentioned above to build the model of Hia<sub>51–166</sub>. CNS<sup>34</sup> and PHENIX.REFINE,<sup>35</sup> together with intermittent manual building in COOT, were used to build the missing residues (residues 973–997) in Hia<sub>937–1098</sub>. Hia residues 937–972 appear to be disordered, hence the electron density map for these residues is not available for model building.

### Structure refinement

The structure models of Hia<sub>51–166</sub>, Hia<sub>307–422</sub>, and Hia<sub>973–1098</sub> were refined by conjugate gradient minimization and simulated annealing implemented in PHENIX.REFINE and CNS with intermittent manual rebuilding, refining individual  $B$ -factors applying TLS correction.<sup>36</sup> The final model of Hia<sub>51–166</sub> ( $R/R_{\text{free}}$  factors=19.4/23.6) contains residues 54–165 and 192 water molecules. The final model of Hia<sub>307–422</sub> ( $R/R_{\text{free}}$  factors=16.2/19.8) contains residues 313–422 and 244 water molecules. The final model of Hia<sub>973–1098</sub> ( $R/R_{\text{free}}$  factors=23.3/28.5) contains residues 973–1098. The detailed structure refinement statistics are reported in Table 2.

### Accession codes

Coordinates of Hia<sub>51–166</sub>, Hia<sub>307–422</sub> and Hia<sub>973–1098</sub> have been deposited to the Protein Data Bank (entry codes 3EMF, 3EMI and 3EMO, respectively).

## Acknowledgements

This work was funded by grant 08087 from the Wellcome Trust (to G.M. and G.W.) and grant RO1-AI44167 from the National Institutes of Health (to J.W.S.). We thank the personnel of beamlines ID14.1, ID14.4, and ID29 at the ESRF for help with data collection.

## Supplementary Data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jmb.2008.09.085](https://doi.org/10.1016/j.jmb.2008.09.085)

## References

1. Turk, D. C. (1984). The pathogenicity of *Haemophilus influenzae*. *J. Med. Microbiol.* **18**, 1–16.
2. Murphy, T. F., Bernstein, J. M., Dryja, D. M., Campagnari, A. A. & Apicella, M. A. (1987). Outer membrane protein and lipooligosaccharide analysis of paired nasopharyngeal and middle ear isolates in otitis media due to nontypeable *Haemophilus influenzae*: pathogenetic and epidemiological observations. *J. Infect. Dis.* **156**, 723–731.
3. Spinola, S. M., Peacock, J., Denny, F. W., Smith, D. L. & Cannon, J. G. (1986). Epidemiology of colonization by nontypeable *Haemophilus influenzae* in children: a longitudinal study. *J. Infect. Dis.* **154**, 100–109.
4. St Geme, J. W., III (2002). Molecular and cellular determinants of non-typeable *Haemophilus influenzae* adherence and invasion. *Cell. Microbiol.* **4**, 191–200.
5. Barenkamp, S. J. & St Geme, J. W., III (1996). Identification of a second family of high-molecular-weight adhesion proteins expressed by non-typeable *Haemophilus influenzae*. *Mol. Microbiol.* **19**, 1215–1223.
6. St Geme, J. W., III, Cutter, D. & Barenkamp, S. J. (1996). Characterization of the genetic locus encoding *Haemophilus influenzae* type b surface fibrils. *J. Bacteriol.* **178**, 6281–6287.
7. Surana, N. K., Cutter, D., Barenkamp, S. J. & St Geme, J. W., III (2004). The *Haemophilus influenzae* Hia autotransporter contains an unusually short trimeric translocator domain. *J. Biol. Chem.* **279**, 14679–14685.
8. Cotter, S. E., Yeo, H. J., Juehne, T. & St Geme, J. W., III (2005). Architecture and adhesive activity of the *Haemophilus influenzae* Hsf adhesin. *J. Bacteriol.* **187**, 4656–4664.
9. Cotter, S. E., Surana, N. K. & St Geme, J. W., III (2005). Trimeric autotransporters: a distinct subfamily of autotransporter proteins. *Trends Microbiol.* **13**, 199–205.
10. Linke, D., Riess, T., Autenrieth, I. B., Lupas, A. & Kempf, V. A. (2006). Trimeric autotransporter adhesins: variable structure, common function. *Trends Microbiol.* **14**, 264–270.
11. Henderson, I. R., Navarro-Garcia, F., Desvaux, M., Fernandez, R. C. & Ala'Aldeen, D. (2004). Type V protein secretion pathway: the autotransporter story. *Microbiol. Mol. Biol. Rev.* **68**, 692–744.
12. Dautin, N. & Bernstein, H. D. (2007). Protein secretion in Gram-negative bacteria via the autotransporter pathway. *Annu. Rev. Microbiol.* **61**, 89–112.
13. Jacob-Dubuisson, F., Fernandez, R. & Coutte, L. (2004). Protein secretion through autotransporter

- and two-partner pathways. *Biochim. Biophys. Acta*. **1694**, 235–257.
14. Roggenkamp, A., Ackermann, N., Jacobi, C. A., Truelzsch, K., Hoffmann, H. & Heesemann, J. (2003). Molecular analysis of transport and oligomerization of the *Yersinia enterocolitica* adhesin YadA. *J. Bacteriol.* **185**, 3735–3744.
  15. Oomen, C. J., van Ulsen, P., van Gelder, P., Feijen, M., Tommassen, J. & Gros, P. (2004). Structure of the translocator domain of a bacterial autotransporter. *EMBO J.* **23**, 1257–1266.
  16. Meng, G., Surana, N. K., St Geme, J. W., III & Waksman, G. (2006). Structure of the outer membrane translocator domain of the *Haemophilus influenzae* Hia trimeric autotransporter. *EMBO J.* **25**, 2297–2304.
  17. Laarmann, S., Cutter, D., Juehne, T., Barenkamp, S. J. & St Geme, J. W. (2002). The *Haemophilus influenzae* Hia autotransporter harbours two adhesive pockets that reside in the passenger domain and recognize the same host cell receptor. *Mol. Microbiol.* **46**, 731–743.
  18. Yeo, H. J., Cotter, S. E., Laarmann, S., Juehne, T., St Geme, J. W., III & Waksman, G. (2004). Structural basis for host recognition by the *Haemophilus influenzae* Hia autotransporter. *EMBO J.* **23**, 1245–1256.
  19. Szczesny, P. & Lupas, A. (2008). Domain annotation of trimeric autotransporter adhesins—daTAA. *Bioinformatics*, **24**, 1251–1256.
  20. Nummelin, H., Merckel, M. C., Leo, J. C., Lankinen, H., Skurnik, M. & Goldman, A. (2004). The *Yersinia* adhesin YadA collagen-binding domain structure is a novel left-handed parallel beta-roll. *EMBO J.* **23**, 701–711.
  21. Hoiczky, E., Roggenkamp, A., Reichenbecher, M., Lupas, A. & Heesemann, J. (2000). Structure and sequence analysis of *Yersinia* YadA and *Moraxella* UspAs reveal a novel class of adhesins. *EMBO J.* **19**, 5989–5999.
  22. Szczesny, P., Lnke, D., Ursinus, A., Bär, K., Schwarz, H., Riess, T. M. *et al.* (2008). Structure of the head of the *Bartonella* adhesin BadA. *PLoS Pathog.* **4**, e1000119.
  23. Luo, Y., Frey, E. A., Pfuetzner, R. A., Creagh, A. L., Knoechel, D. G., Haynes, C. A. *et al.* (2000). Crystal structure of enteropathogenic *Escherichia coli* intimin–receptor complex. *Nature*, **405**, 1073–1077.
  24. Sauer, F. G., Futterer, K., Pinkner, J. S., Dodson, K. W., Hultgren, S. J. & Waksman, G. (1999). Structural basis of chaperone function and pilus biogenesis. *Science*, **285**, 1058–1061.
  25. Choudhury, D., Thompson, A., Stojanoff, V., Langermann, S., Pinkner, J., Hultgren, S. J. & Knight, S. D. (1999). X-ray structure of the FimC–FimH chaperone–adhesin complex from uropathogenic *Escherichia coli*. *Science*, **285**, 1061–1066.
  26. Krojer, T., Sawa, J., Schafer, E., Saibil, H. R., Ehrmann, M. & Clausen, T. (2008). Structural basis for the regulated protease and chaperone function of DegP. *Nature*, **453**, 885–890.
  27. Eppens, E. F., Nouwen, N. & Tommassen, J. (1997). Folding of a bacterial outer membrane protein during passage through the periplasm. *EMBO J.* **16**, 4295–4301.
  28. Ieva, R., Skillman, K. M. & Bernstein, H. D. (2008). Incorporation of a polypeptide segment into the beta-domain pore during the assembly of a bacterial autotransporter. *Mol. Microbiol.* **67**, 188–201.
  29. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
  30. CCP4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Sect. D*, **50**, 760–763.
  31. Sheldrick, G. M. & Schneider, T. R. (1997). SHELXL: high-resolution refinement. *Methods Enzymol.* **277**, 319–343.
  32. de laFortelle, E. & Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494.
  33. Lamzin, V. S. & Wilson, K. S. (1993). Automated refinement of protein models. *Acta Crystallogr. Sect. D*, **49**, 129–147.
  34. Brunger, A. T. & Rice, L. M. (1997). Crystallographic refinement by simulated annealing: methods and applications. *Methods Enzymol.* **277**, 243–269.
  35. Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). CCP4 Newsl. **42**; contribution 8.
  36. Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. Sect. D*, **57**, 122–133.
  37. Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Crystallogr.* **A47**, 392–400.
  38. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291.